Education Delays Dementia Onset and Mitigates Genetic Risk for Alzheimer's Disease

Silvia H. Barcellos (USC and NBER)
Leandro S. Carvalho (USC)
Ken Langa (University of Michigan)
Sneha Nimmagadda (USC)
Patrick Turley (USC)

Abstract: We combine a natural experiment with data from the UK Biobank to study the causal effect of education on risk for Alzheimer's Disease and Related Dementias (ADRD). The natural experiment generated as-good-as-random variation in secondary education. ADRD was ascertained based on medical history, hospital records, and death registries, which are not biased by selective sample attrition. We find evidence that education reduces ADRD risk and delays its onset. The reduction in ADRD caused by education is larger for individuals with greater genetic risk of AD, implying that this genetic risk is not immutable and can be modified by social policy.

Introduction

Gains in life expectancy and population aging are driving a sharp rise in Alzheimer's Disease and Related Dementias (ADRD), with predicted global cases of 131.5 million people by 2050 (Prince et al. 2015, Livingston et al. 2017). In addition to imposing health and emotional costs on ADRD patients and their families, the severe cognitive and functional impairment that characterizes ADRD has a large economic cost on society. In the U.S. alone, ADRD's cost is estimated at $305 billion (Wong 2020). Without effective interventions to prevent or delay its onset, the economic burden of dementia could grow by as much as three-fold in the next 35 years (Livingston et al. 2017). For these reasons, it is crucial to understand what factors and interventions, if any, can decrease ADRD risk and delay cognitive and functional decline in older ages.

Lower educational attainment has been identified as a major risk factor for ADRD, as more education is strongly associated with better later-life cognition and lower ADRD prevalence (Sharp and Gatz 2011, Maccora et al. 2020). However, little is known about whether and how much of this association reflects a causal effect from education to ADRD. Part of the challenge in distinguishing correlation from causation is that third factors, such as childhood circumstances, socioeconomic background, and genetics, may confound the relationship between education and

ADRD. Also, there may be individual characteristics, such as self-control, that drive both behavioral risk factors and educational attainment. Establishing whether education has a causal effect on late-life ADRD risk is essential for identifying effective clinical and policy interventions.

Another obstacle to studying the education-ADRD relationship is that dementia is difficult to diagnose and measure. Cognitive decline often progresses slowly, and the threshold of disability that defines ADRD can be hard to identify reliably. Moreover, while brain pathology and cognition are correlated, there is not a one-to-one correspondence between them.[18] Selective sample attrition is also a concern for most survey-based measures of ADRD (Weir, Faul, and Langa 2011): Survey participants who develop dementia are more likely to drop out, either because of survival bias or because they can no longer answer survey questions reliably.

We overcome these challenges by combining a natural experiment with dementia data constructed from administrative health records for a large aging cohort. The natural experiment, a well-known compulsory schooling age reform in the United Kingdom, generated as-good-as-random variation in secondary education, allowing us to estimate the causal effect of education on dementia.

We use data from the UK Biobank (UKB). The UKB's Outcome Adjudication Group has identified dementia outcomes using hospital records, mortality data, and (self-reported) medical history. Date of first dementia diagnosis is also recorded, allowing us to study effects on age at onset. A validation study shows that at least 84.5% of the diagnoses coded based on this algorithm are true positives (Wilkinson et al. 2019). Crucially, these data are not biased by selective attrition; the study continues to follow its participants through linkages to administrative health and death registries.

The UKB also genotyped its participants, which allows us to investigate whether education can mitigate the genetic risk for ADRD (Cook and Fletcher 2015). ADRD has a strong genetic basis, with an estimated twin-based heritability of AD as high as 80% (Gatz et al. 2006, Bird 2005, Breteler 2000) and a SNP-heritability of 33% (Ridge et al. 2013). Individuals who inherit certain variants are at much greater risk of developing ADRD—having two copies of APOE4 is associated with a 10-fold higher risk, for example. While it is currently easy to measure one's genetic risk, it is hard to know how to react to this information. Whether behavioral changes, clinical interventions or social policies (such as educational reforms) can mitigate such genetic risks is an open question.

While previous studies found evidence that education causally improves cognition and slows cognitive decline (e.g., Banks and Mazonna 2012; Nguyen et al. 2016), we are aware of only one other study that has examined the causal relationship of education with ADRD risk (Seblova et al. 2021). They exploited temporal variation in when school districts in Sweden implemented a compulsory schooling age reform; early-adopters implemented the reform as early as 1936 while late-adopters implemented as late as 1953. One concern is that students in early-adopter and late-adopter school districts may have had different ADRD risk even in the absence of the reform. The paper finds no effect of education on ADRD risk; the reform also had limited effects on socioeconomic outcomes such as income.

In contrast, we study a compulsory schooling age reform that lends itself to a credible regression discontinuity design. Multiple studies have used this reform before, speaking to the widely-accepted credibility of this natural experiment (Clark and Royer 2013, Davies et al. 2018, Barcellos et al. 2018). Importantly, this reform did affect socioeconomic status, including income.

The main contribution of this paper is to provide rigorous evidence that education can reduce ADRD risk and delay its onset. Another is to combine the natural experiment with genetic data. We find that the reduction in risk was larger for those at greater genetic risk of developing ADRD. These results are encouraging, suggesting that education can indeed mitigate the genetic risk for ADRD.

The Natural Experiment

In 1972, England, Scotland, and Wales increased the minimum age at which students could drop out of school from 15 to 16 years of age. The reform (The 1972 Raising of School-Leaving Age Act or ROSLA) affected only students born on or after September 1, 1957, who had to stay in school until at least age 16. Students born on August 31, 1957 or before were still allowed to drop out at age 15. Since September 1, 1957 is an arbitrary date, students born right before and right after this date would have similar ADRD risk in the absence of this policy. We exploit this natural experiment using a regression discontinuity design.

The 1972 ROSLA introduced a sharp discontinuity in the relationship between education and date of birth. Figure 1A shows the fraction who stayed in school until age 16 jumps for those born after September 1, 1957. They are 14 percentage points more likely to stay in school until age 16 than their counterparts (see column (1) of Table 1).

This setting provides a unique opportunity to study the causal effects of education on ADRD. If there is a causal relationship between education and ADRD, as hypothesized, then one would expect to observe a corresponding discontinuity in the relationship between ADRD and date of birth at September 1, 1957. Individuals born right before and right after this date would have had similar lifetime ADRD risk in the absence of the policy. Therefore, any observed discontinuity can be attributed to the difference in schooling engendered by the policy.

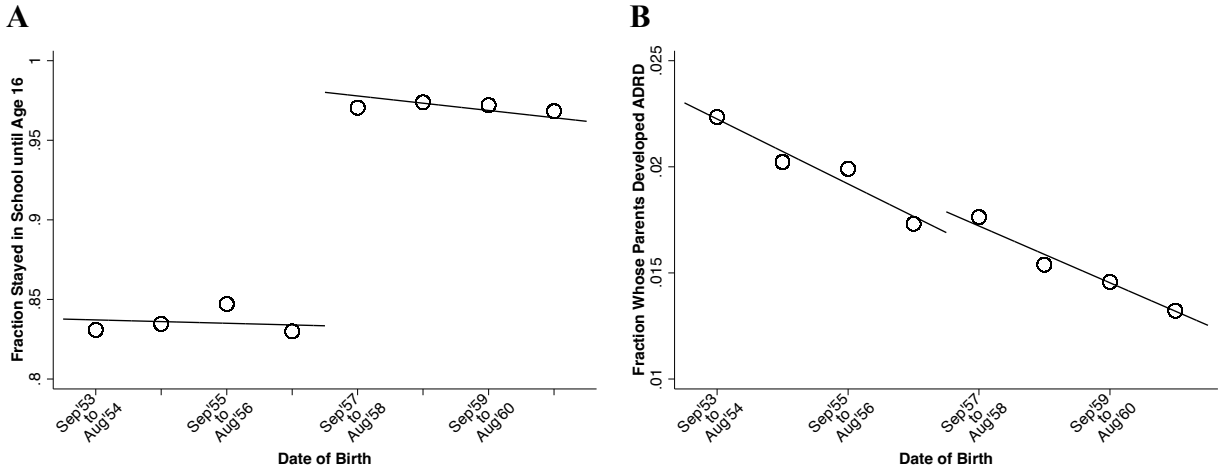


Fig. 1: The Natural Experiment. The figures plot (A) the fraction of each birth cohort who stayed in school until age 16 and (B) the fraction of each birth cohort whose parents developed ADRD. The cohort born between September 1, 1957 and August 31, 1958 was the first affected by the reform. The lines show linear trends in date of birth, which are allowed to be different for those born before and after September 1, 1957. Figure (A) shows that the reform generated a discontinuity in the relationship between education and date of birth. Figure (B) shows that *the parents* of those born right after September 1, 1957 are as likely to have ADRD as *the parents* of those born right before September 1, 1957. This illustrates that the latter provides a valid counterfactual of what would have happened to the former had they not been forced to stay in school until age 16.

For the natural experiment to be valid, the two groups would have had to be comparable before the reform, including in terms of their family medical history and their genetics. In particular, there should not be any *pre-reform* factor that changed discontinuously at September 1, 1957. Figure 1B provides evidence of the validity of the counterfactual. It shows prevalence rates of ADRD *among the parents of study participants*. Parents of those born right after September 1,

1957 are as likely to suffer from ADRD as *the parents* of those born right before September 1, 1957. The Supplementary Materials also show that those born before and after September 1, 1957 are comparable in terms of their genetics, including in terms of their genetic risks for ADRD, and other pre-reform characteristics. This provides strong evidence that the ROSLA did induce a valid natural experiment.

Education Reduces Risk of ADRD and Delays Onset

Our findings indicate that education has a causal negative effect on ADRD risk. Figure 2A shows a sharp decline in ADRD risk among those born after September 1, 1957. They are 0.2 percentage points less likely to have been diagnosed with dementia (p-value of 0.01) – see column (2) of Table 1. This is a large decline considering that risk among those born before September 1, 1957 was approximately 0.48%.

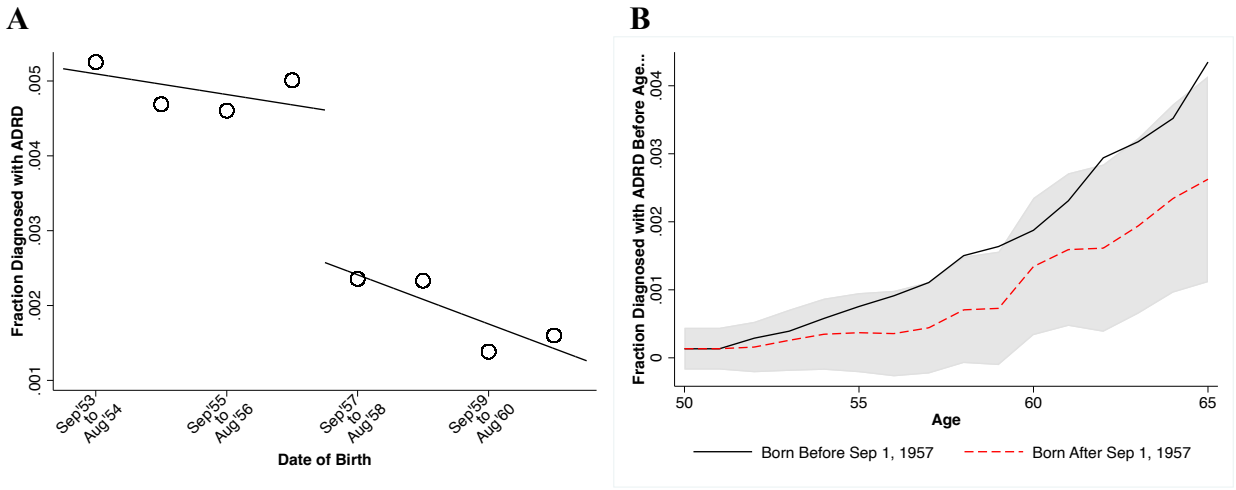


Fig. 2: Education Reduces ADRD Risk and Delays Onset. Figure (A) plots the fraction of each birth cohort whose was diagnosed with dementia. The cohort born between September 1, 1957 and August 31, 1958 was the first affected by the reform. The lines show linear trends in date of birth, which are allowed to be different for those born before and after September 1, 1957. Figure (A) shows a sharp decline in ADRD prevalence rate among those born after September 1, 1957, illustrating the causal effects of education on ADRD. Figure (A) shows the effect on prevalence rates as of December of 2022, when individuals born on September 1, 1957 were 65 years old. Figure (B) shows the effect on risk at other ages. It plots hazard rates of dementia diagnosis, separately for those born right before September 1, 1957 (solid black line) and for those born right after September 1, 1957 (red dashed line). The shaded area shows 95% confidence intervals for the difference between the two hazards; the difference is statistically significant at the 5% level when the solid black line is outside the shaded area. The figure suggests that education may delay the onset of ADRD.

Column (3) of Table 1 estimates that staying in school one year longer reduces ADRD risk by 1.4 percentage points (p-value of 0.036). This effect is estimated through two-stage least squares, where the indicator for being born after September 1, 1957 is used to instrument for school-leaving age. While this point estimate is implausibly large, the confidence interval implies that staying in school a year longer reduces ADRD risk by at least 0.09 percentage points.

To increase statistical power, columns (4) to (6) of Table 1 restrict the sample to individuals who left school at age 18 or younger. This sample restriction is valid because, as shown in the Supplementary Materials, the reform did not affect the fraction of students who stayed in school past age 18. Column (6) estimates that staying in school one year longer reduces ADRD prevalence by approximately 1 percentage point (p-value of 0.01). The upper bound of the 95% confidence interval is a reduction of 0.23 percentage points.

Figure 2A and Table 1 show the effect on ADRD risk as of December of 2022, when individuals born on September 1, 1957 were 65 years old. Figure 2B shows the effect on risk at other ages. It plots hazard rates of dementia diagnosis, separately for those born right before September 1, 1957 (solid black line) and for those born right after (red dashed line). Intuitively, the vertical distance between the red and black lines at a given age corresponds to the effect of the policy on ADRD risk at the given age. The shaded area shows 95% confidence intervals for the difference between the two hazards—the difference is statistically significant at the 5% level when the solid black line is outside the shaded area. The figure suggests that the additional schooling induced by the 1972 ROSLA have delayed ADRD onset. We formally test this hypothesis using a Cox proportional hazard model (see Supplementary material), and we can reject the hypothesis that the two groups share the same hazard rate (p-value of 0.025). Given these results, we anticipate that the gap between the two groups will widen over time as this birth cohort ages further.

Table 1: Regression Estimates of Effect of The Reform and of The Effect of Education. The dependent variables label each column. ADRD prevalence is measured in percentage points. "Born after" is an indicator for being born on or after September 1, 1957. All regressions include linear trends in date of birth; those born before and after September 1, 1957 are allowed to have different trends. Columns (1), (2), (4), and (5) are OLS estimates. Columns (3) and (6) are estimated through two-stage least squares, where "Born after" is used to instrument for school-leaving age. In columns (4) to (6), the sample is restricted to individuals who dropped out at age 18 or younger. Regressions in Panel B include either an interaction of the genetic risk index with "Born after" (in columns (1), (2), (4), and (5)) or an interaction of the index with school-leaving age (in columns (3) and (6)); these regressions also control for the index. The ADRD genetic risk is standardized to have a standard deviation of one. The last row in each panel shows the mean of the dependent variable among those who were born before September 1, 1957. Robust standard errors between parentheses. In Panel A, the number of observations is 89,259 for columns (1) to (3) and 49,457 for columns (4) to (6). In Panel B, the number of observations is 82,250 for columns (1) to (3) and 45,715 for columns (4) to (6). The number of observations is smaller in Panel B because the genetic risk index is not available for some participants.

	All			**Dropped Out at Age 18 or Younger**		
	School-Leaving Age (1)	*ADRD (in p.p.)* (2)	*ADRD (in p.p.)* (3)	*School-Leaving Age* (4)	*ADRD (in p.p.)* (5)	*ADRD (in p.p.)* (6)
PANEL A						
Born after	0.14	-0.20		0.32	-0.31	
	(0.04)	(0.08)		(0.02)	(0.12)	
School-Leaving Age			-1.44			-0.96
			(0.69)			(0.37)
Mean of Y among Those Born before	18.71	0.48	0.48	16.17	0.60	0.60
PANEL B						
Born after × Genetic Risk	0.004	-0.18		-0.01	-0.19	
	(0.02)	(0.06)		(0.01)	(0.08)	
School-Leaving Age × Genetic Risk			-1.38			-0.70
			(1.13)			(0.31)
Born after	0.14	-0.19		0.32	-0.29	
	(0.04)	(0.08)		(0.02)	(0.12)	
School-Leaving Age			-1.68			-0.97
			(0.94)			(0.38)
Genetic Risk	0.01	0.24	5.39	0.01	0.24	1.08
	(0.01)	(0.05)	(4.28)	(0.01)	(0.07)	(0.44)
Mean of Y among Those Born before	18.69	0.47	0.47	16.16	0.58	0.58

Reduction in ADRD Risk Is Larger for Those with Greater ADRD Genetic Risk

A common perception is that the influence of genetics on a disease such as ADRD cannot be altered by environmental interventions – presumably because it is assumed that this influence reflects a biological, immutable relationship. We proceed to investigate whether the 1972 ROSLA changed the relationship between genetics and ADRD. As a measure of one's genetic risk of developing ADRD, we construct an AD polygenic index that includes both a polygenic component and a person's APOE genotype (see Supplemental Materials for details).

Figure 3 investigates how two groups – those in the bottom and in the top halves of the distribution of this index – were affected by the ROSLA. Because APOE has such a large effect on ADRD risk, most of this variation is driven by APOE (58% of those in the top half have at least one E4 allele, while only 0.02% of those in the bottom half do). The black markers and black line show effects on those in the bottom half with lower ADRD genetic risk. The red markers and red line show effects on those in the top half with greater ADRD genetic risk.

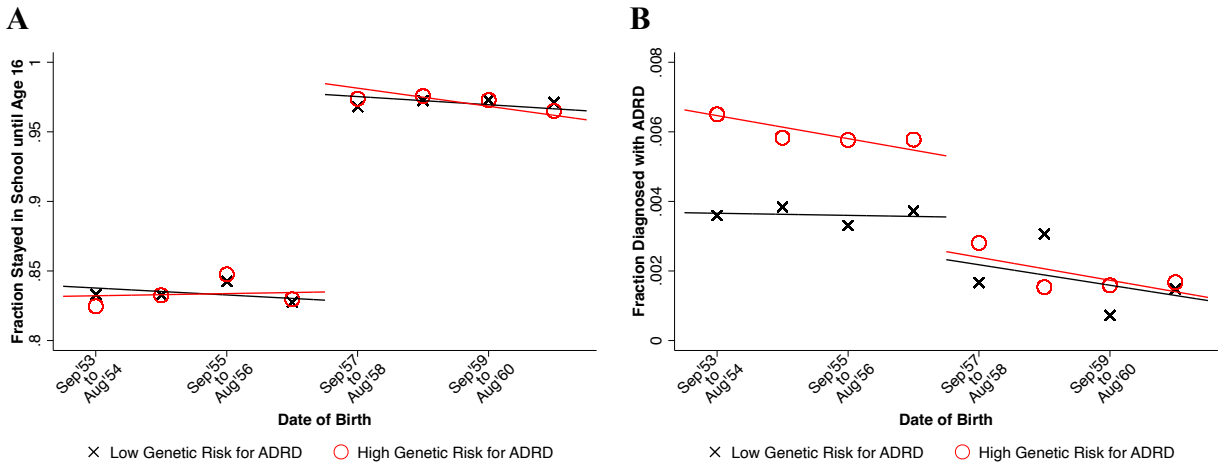


Fig. 3: Reduction in ADRD Risk Is Larger for Those with Greater ADRD Genetic Risk. We measured one's genetic risk of developing ADRD by constructing an ADRD polygenic index that includes both a polygenic component and a person's APOE genotype. Separate results are shown for those in the bottom half (black) and in the top half (red) of the distribution of this index. The figures plot (A) the fraction of each birth cohort who stayed in school until age 16 and (B) the fraction of each birth cohort who was diagnosed with ADRD. The cohort born between September 1, 1957 and August 31, 1958 was the first affected by the reform. The lines show linear trends in date of birth, which are allowed to be different for those born before and after September 1, 1957.

Figure (A) shows that the reform affected the education of the two groups similarly. Figure (B) shows that it had a larger effect on the prevalence rate of those in the top half.

Figure 3A shows the policy affected the education of the two groups similarly. In both cases, those born right after September 1, 1957 are 14 percentage points more likely to stay in school until age 16. Thus, any differences in the effect of the policy on ADRD across the two groups cannot be explained by differences in how much the policy impacted the education of these groups.

Figure 3B shows ADRD prevalence rates for the two groups. For the cohorts born before September 1, 1957, there is a gap between the red and black lines: prevalence was about 0.22 percentage points higher for individuals in the top half of the distribution of the genetic risk index. If genetic risk were immutable, the gap between the red and black lines would remain the same for those born after September 1, 1957.

The figure shows, however, that the additional schooling caused a greater reduction in ADRD risk for those in the top half: a reduction of 0.29 percentage points (p-value of 0.001), from 0.56% to 0.26%. In contrast, it reduced the prevalence among those in the bottom half by approximately 0.09 percentage points (p-value of 0.301), from 0.33% to 0.24%. Visually, it is clear that the policy closed virtually the entire gap between the two groups.

Panel B of Table 1 confirm these results. We re-estimate the same specifications estimated in Panel A, but now allowing the effects to vary with one's continuous ADRD genetic risk. The regressions include either an interaction of the genetic risk index with the indicator for being born after September 1, 1957 (in columns (1), (2), (4), and (5)) or an interaction of the index with school-leaving age (in columns (3) and (6)). The regressions also control for the index, which is standardized to have a standard deviation of one. Consistent with Figure 3A, column (1) in Panel B shows that the effect of the policy on education was the same for high- and low-risk individuals.

Nevertheless, the interaction term in column (2) shows that the reduction in ADRD prevalence was larger among individuals with higher genetic risk, as seen in Figure 3B. For those born before September 1, 1957, a one standard deviation increase in the genetic risk index was associated with a 0.24 percentage points increase in prevalence. In contrast, one cannot reject the hypothesis that, among those born after September 1, 1957, there is no association between the index and ADRD prevalence. Column (6) indicates that education provides greater protection for those at greater risk: staying in school a year longer reduces prevalence by 1 percentage point for those with

average risk and by 1.7 percentage points for those with a genetic risk one standard deviation above the mean. These results are encouraging because they suggest that ADRD generic risk can be mitigated, and that changes in environmental circumstances, such as education, can reduce ADRD prevalence even among individuals with high genetic risk.

Examining Potential Mediators: Income and Risk Factors

Figure 4A examines some potential channels through which the additional schooling may have affected ADRD risk in later in life. It shows intent-to-treat estimates of the effect of the policy on the outcomes listed in the rows. Those affected by the 1972 ROSLA are more likely to have an income of £31,000 or more (p-value 0.001) and less likely to have diabetes (p-value 0.045) and to have had a stroke or myocardial infarction (p-value 0.045). The point estimates also indicate a reduction in obesity of 0.8 percentage points, but this effect is not statistically significant (p-value of 0.139). Figure 4B re-estimates the effects on these risk factors, separately for those in the bottom and in the top halves of the distribution of the ADRD genetic risk index. Interestingly, the effects on diabetes and obesity seem to be driven almost exclusively by the high-genetic-risk group.

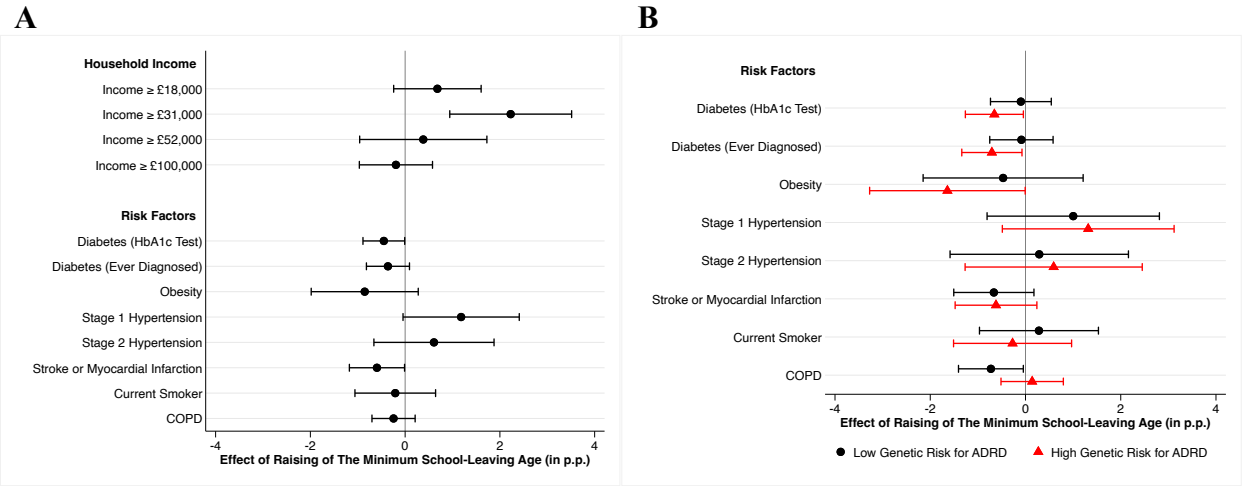


Fig. 4: Potential Mediators: Income and Risk Factors. The figure examines some potential channels through which the additional schooling may have affected ADRD risk. It shows intent-to-treat estimates of the effect of the policy on the outcomes listed in the rows. Estimates are in percentage point units. The top panel of each figure estimates the effects on annual household income. The bottom panel estimates effects on various risk factors for ADRD. While (A) shows estimates for the entire sample, (B) estimates separate effects on those in the bottom half (black) and on those in the top half (red) of the distribution of the ADRD genetic risk index. The brackets show 95% confidence intervals. Robust standard errors.

Discussion

Scholars have noted the need for high-quality evidence on which factors can reduce ADRD risk and delay its onset. Characteristics that are highly associated with ADRD risk have been deemed as risk factors for dementia. However, there is limited evidence as to whether these characteristics actually causally affect ADRD. Such evidence is paramount for designing policies that can prevent and effectively treat dementia.

Education is a case in point. While lower education has been singled out as a major risk factor for ADRD, there is limited evidence on the causal effect of education on dementia. We are aware of just one other paper that studies this causal relationship using a natural experiment combined with dementia measures derived from administrative health records. Seblova et al. (2021) study the effects of a compulsory schooling age reform in Sweden. Inpatient and death registries were used to measure dementia diagnosis for over 1.3 million individuals. They find no effect of the reform on ADRD risk. There are some concerns about the credibility of the paper's empirical strategy, which exploits the staggered rollout of the reform over an extended period of time of 17 years.

In contrast, we studied a natural experiment with widely-accepted credibility and that lends itself to a compelling regression discontinuity design. We find unequivocal evidence that the additional schooling generated by the 1972 ROSLA reduced ADRD risk and delayed ADRD onset. These results indicate that ADRD risk can be modified by policy, a welcome finding given the expected exponential growth in cases of dementia and their large costs. To our knowledge, this is the first paper to document a causal relationship between education and ADRD. Encouragingly, we find that the reduction in ADRD risk was larger for those at greater genetic risk of developing ADRD – rejecting the common perception that the influence of genetics on a disease such as ADRD cannot be altered by environmental interventions.

To translate this evidence into policy-relevant interventions, it is important to understand the channels through which education affects ADRD risk. While the reform in Sweden studied by Seblova et al. (2021) seems to have had limited, if any, effects on socioeconomic outcomes, it is clear that the 1972 ROSLA improved socioeconomic status – *suggesting* that this may be one of the primary mechanisms through which education reduces ADRD risk. The ROSLA may have

also reduced some negative health outcomes, such as diabetes, obesity, stroke and infarction, but much remains to be understood.

This study comes with several limitations. First, our results are specific to the lower-education population in the UK affected by the ROSLA. Admittedly, the effects of education on ADRD may vary with the level of education and across countries. Furthermore, the cohort studied in this paper is relatively young; those born in September 1, 1957 were 65 years old in 2022. Although the ADRD prevalence in our sample is lower than 1%, we are still able to identify a clear effect of education on ADRD, which suggests that even larger effects will be seen as this cohort ages. Follow-up with this cohort will be important.

REFERENCES

Banks, J. & Mazzonna, F. The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. *Econ. J.* 122, 418–448 (2012).

Barcellos, S. H., Carvalho, L. S. & Turley, P. Education can reduce health differences related to genetic risk of obesity. *Proc. Natl. Acad. Sci.* **115**, E9765 (2018).

Bird, T. D. Genetic factors in Alzheimer's disease. *N. Engl. J. Med.* **352**, 862–864 (2005).

Breteler, M. M. B. Vascular risk factors for Alzheimer's disease:: An epidemiologic perspective. *Neurobiol. Aging* **21**, 153–160 (2000).

Clark, D. & Royer, H. The Effect of Education on Adult Mortality and Health: Evidence from Britain. *Am. Econ. Rev.* **103**, 2087–2120 (2013).

Cook, C. Justin, and Jason M. Fletcher. "Can education rescue genetic liability for cognitive decline?." *Social Science & Medicine* 127 (2015): 159-170.

Davies, N. M., Dickson, M., Smith, G. D., Van Den Berg, G. J. & Windmeijer, F. The causal effects of education on health outcomes in the UK Biobank. *Nat. Hum. Behav.* (2018). doi:10.1038/s41562-017-0279-y

Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).

Livingston, G. *et al.* Dementia prevention, intervention, and care. *Lancet* **390**, 2673–2734 (2017).

Maccora, J., Peters, R. & Anstey, K. J. What does (low) education mean in terms of dementia risk?: a systematic review and meta-analysis highlighting inconsistency in measuring and operationalising education. *SSM-Population Heal.* 100654 (2020).

Nguyen, T. T. *et al.* Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. *Ann. Epidemiol.* **26**, 71–76 (2016).

Prince, M. J. *et al.* World Alzheimer Report 2015-The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends. (2015).

Ridge, P. G., Mukherjee, S., Crane, P. K. & Kauwe, J. S. K. Alzheimer's disease: Analyzing the missing heritability. *PLoS One* (2013). doi:10.1371/journal.pone.0079771

Seblova, D., Fischer, M., Fors, S., Johnell, K., Karlsson, M., Nilsson, T., ... & Lager, A. (2021). Does prolonged education causally affect dementia risk when adult socioeconomic status is not altered? A Swedish natural experiment in 1.3 million individuals. *American Journal of Epidemiology*, *190*(5), 817-826.

Sharp, E. S. & Gatz, M. The relationship between education and dementia an updated systematic review. *Alzheimer Dis. Assoc. Disord.* **25**, 289 (2011).

Weir D, Faul J, Langa KM. Proxy interviews and bias in cognition measures due to non-response in longitudinal studies: a comparison of HRS and ELSA. Longitudinal and Life-course Studies. 2011; 2(2): 170-184.

Wilkinson, Tim, Christian Schnier, Kathryn Bush, Kristiina Rannikmäe, David E. Henshall, Chris Lerpiniere, Naomi E. Allen et al. "Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data." *European journal of epidemiology* 34 (2019): 557-565.

Wong, W. Economic burden of Alzheimer disease and managed care considerations. *Am. J. Manag. Care* **26**, S177–S183 (2020).